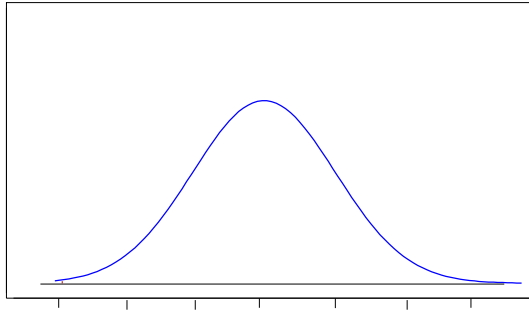


Practice First Midterm Exam

Honors Stat 1000 Fall 2012 (Pfenning)

This is a closed book exam worth 150 points. You are allowed to use a calculator and a two-sided sheet of notes. There are 8 problems, with point values as shown. If you want to receive partial credit for wrong answers, show your work. Don't spend too much time on any one problem.

1. (20 pts.) Weight gains during pregnancy were found to be approximately normally distributed with mean 34 pounds, standard deviation 13 pounds.
 - (a) Use the 68-95-99.7 Rule to fill in numbers for the seven indicated points of the horizontal axis on this curve showing the distribution of weight gains.



- (b) Give a range for the percentage of weight gains that were more than 50 pounds: between _____% and _____%
- (c) How much did a woman gain, if her z score is -2.3 ? _____
- (d) In reality, the normal curve sketched above as a model for women's weight gains is
 - i. probably quite accurate
 - ii. probably somewhat inaccurate: the actual curve would be left-skewed
 - iii. probably somewhat inaccurate: the actual curve would be right-skewed
 - iv. not just "probably inaccurate", but impossible

2. (20 pts.) Identify the variables in each of the following situations as one of these:

- i single categorical
- ii single quantitative
- iii two categorical
- iv categorical explanatory and quantitative response
- v quantitative explanatory and categorical response
- vi two quantitative

(a) SAT is a de facto IQ test

- (i) C (ii) Q (iii) $C \rightarrow C$ (iv) $C \rightarrow Q$ (v) $Q \rightarrow C$ (vi) $Q \rightarrow Q$

(b) In 1990, 11 percent of survey respondents in Olmsted County believed that blacks are more violent than whites. In 2007, 30 percent believed so.

- (i) C (ii) Q (iii) $C \rightarrow C$ (iv) $C \rightarrow Q$ (v) $Q \rightarrow C$ (vi) $Q \rightarrow Q$

(c) Can a woman's score on a psychological test for self-objectification predict whether or not she will refrain from eating the second Twix bar after she's tried on a swimsuit?

- (i) C (ii) Q (iii) $C \rightarrow C$ (iv) $C \rightarrow Q$ (v) $Q \rightarrow C$ (vi) $Q \rightarrow Q$

(d) High school students are less likely than adults to be very interested in economics.

- (i) C (ii) Q (iii) $C \rightarrow C$ (iv) $C \rightarrow Q$ (v) $Q \rightarrow C$ (vi) $Q \rightarrow Q$

(e) 37.5% of survey respondents reported their average number of "NATURAL" bowel movements to be "1 to 2 bowel movements per day."

- (i) C (ii) Q (iii) $C \rightarrow C$ (iv) $C \rightarrow Q$ (v) $Q \rightarrow C$ (vi) $Q \rightarrow Q$

(f) Students calculated their z-scores for the number of minutes they spent doing homework the day before.

- (i) C (ii) Q (iii) $C \rightarrow C$ (iv) $C \rightarrow Q$ (v) $Q \rightarrow C$ (vi) $Q \rightarrow Q$

(g) How does the amount of isoproterenol received by a subject affect his or her heart rate?

- (i) C (ii) Q (iii) $C \rightarrow C$ (iv) $C \rightarrow Q$ (v) $Q \rightarrow C$ (vi) $Q \rightarrow Q$

(h) Do Pittsburgh, Detroit, and Atlanta differ in terms of how common it is for realtors to treat black renters adversely?

- (i) C (ii) Q (iii) $C \rightarrow C$ (iv) $C \rightarrow Q$ (v) $Q \rightarrow C$ (vi) $Q \rightarrow Q$

(i) Is the rate of forest depletion in a country related to what continent it's in?

- (i) C (ii) Q (iii) $C \rightarrow C$ (iv) $C \rightarrow Q$ (v) $Q \rightarrow C$ (vi) $Q \rightarrow Q$

3. (20 pts.) The Pittsburgh Post Gazette reports: “The gender gap has widened when it comes to hygiene, according to the latest stakeout by the ‘hand washing police.’ One-third of men didn’t bother to wash after using the bathroom, compared with 12 percent of women, said the researchers who spy on people in public restrooms...Two years ago, the last time the survey was done, only one-quarter of men didn’t wash, compared with 10 percent of women...The latest study was based on observations last month of more than 6,000 people in four cities.”

- (a) Which **two** of these are taken to be explanatory variables?
(i) gender (ii) washing hands (iii) city (iv) year (v) month (vi) number of people
- (b) For the purpose of this study, researchers most likely positioned themselves in the designated restrooms and observed a group of individuals who used the facilities over a certain period of time. What words best describe the two main features of their method? _____sampling, _____sampling
- (c) During that two-year period, the rate of failing to wash hands for men
(i) increased by about $\frac{1}{5}$ (ii) increased by about $\frac{1}{4}$ (iii) increased by about $\frac{1}{3}$
(iv) increased by about $\frac{1}{2}$ (v) decreased by about $\frac{1}{5}$ (vi) decreased by about $\frac{1}{4}$
(vii) decreased by about $\frac{1}{3}$ (viii) decreased by about $\frac{1}{2}$
- (d) During that two-year period, the gap between percentages failing to wash hands for men and women
(i) increased by about $\frac{1}{5}$ (ii) increased by about $\frac{1}{4}$ (iii) increased by about $\frac{1}{3}$
(iv) increased by about $\frac{1}{2}$ (v) decreased by about $\frac{1}{5}$ (vi) decreased by about $\frac{1}{4}$
(vii) decreased by about $\frac{1}{3}$ (viii) decreased by about $\frac{1}{2}$

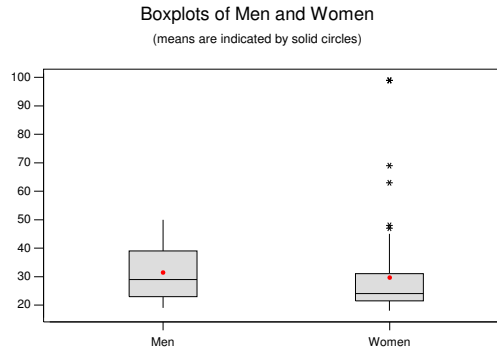
4. (5 pts.) **Case 1:** You are informed that a poker hand of 5 cards, randomly dealt from 52, contains an ace.

Case 2: You are informed that a poker hand of 5 cards, randomly dealt from 52, contains the ace of clubs.

In which case is it more likely that the poker hand includes at least one more ace?

- (i) Case 1 (ii) Case 2 (iii) equally likely (iv) not enough information

5. (25 pts.) Here are side-by-side boxplots for ages of men-seeking-women and women-seeking-men advertised on Pittsburgh's Craig's List.



- (a) This is a (i) paired (ii) two-sample (iii) several sample study
- (b) For which of these would 9 be a reasonable guess? (You can circle anywhere from none to all six.)
- i. minimum value for men
 - ii. minimum value for women
 - iii. IQR for men
 - iv. IQR for women
 - v. standard deviation for men
 - vi. standard deviation for women
- (c) The shapes are best described as
- i. both skewed left (or low outliers)
 - ii. men skewed left (or low outliers) and women skewed right (or high outliers)
 - iii. men skewed right (or high outliers) and women skewed left (or low outliers)
 - iv. both skewed right (or high outliers)
- (d) Apparently, $Q3 + 1.5(IQR)$ for the women is approximately _____ (estimate to the nearest 5 years)
- (e) Suppose we'd like to use the data to draw conclusions about the mean age of all Pittsburgh singles who are seeking a partner. One problem is that people tend to lie about their age. This suggests we may have (i) a non-representative sample (ii) an inaccurate assessment of the sampled values

6. (25 pts.) A study about weight-gain during pregnancy included background information on race and marital status. A two-way table of the data can be used to explore the relationship between these two variables.

	Married	Not Married	Total
Caucasian	2641	589	3230
Asian	513	57	570
Total	3154	646	3800

- (a) The most relevant conditional proportions to compare are (i) $\frac{2641}{3154}$ vs. $\frac{589}{646}$
(ii) $\frac{2641}{3800}$ vs. $\frac{589}{3800}$ (iii) $\frac{2641}{3230}$ vs. $\frac{513}{570}$ (iv) $\frac{2641}{3800}$ vs. $\frac{513}{3800}$ (v) $\frac{3154}{3800}$ vs. $\frac{646}{3800}$ (vi) $\frac{3230}{3800}$ vs. $\frac{570}{3800}$
- (b) The difference between the relevant conditional percentages is about _____ percent (round to the nearest whole number).
- (c) Find the overall proportion who were married. _____
- (d) Explain why the overall proportion is *not* close to being halfway between the two proportions in (a).
- (e) If there were equal proportions married for Caucasians and Asians, about how many of the 570 Asians would be married? (round to the nearest whole number)

7. (5 pts.) Suppose the scatterplot for a relationship between two variables consists of points occurring exactly on a horizontal line. Discuss the strength of the relationship. (There are a variety of approaches to take that require no formula. Or, if you like, you can refer to the formula for slope of the regression line: $b_1 = r \frac{s_y}{s_x}$ or the formula for correlation $r = \frac{1}{n-1} \left[\left(\frac{x_1 - \bar{x}}{s_x} \right) \left(\frac{y_1 - \bar{y}}{s_y} \right) + \dots + \left(\frac{x_n - \bar{x}}{s_x} \right) \left(\frac{y_n - \bar{y}}{s_y} \right) \right]$)

8. (30 pts.) This scatterplot shows average commercial insurance payments, in thousands of dollars, made to ten Pittsburgh-area hospitals for heart valve surgery vs. coronary bypass surgery.

- (a) What is the correlation r ? _____
- (b) UPMC averaged 35 thousand for coronary bypass surgery and 54 thousand for valve surgery. If it were omitted from the regression, the correlation would
 - (i) increase (ii) decrease (iii) stay the same (iv) there's no way of knowing in advance
- (c) UPMC averaged 35 thousand for coronary bypass surgery and 54 thousand for valve surgery. It could be considered an
 - (i) outlier (ii) influential observation (iii) both (iv) neither
- (d) If we reversed roles, taking valve surgery as the explanatory variable and bypass surgery as the response, which of these would change?
 - (i) equation of the regression line (ii) correlation r (iii) both (iv) neither
- (e) Bypass surgery averaged 24 thousand dollars at Allegheny General. Use the regression line to predict its valve surgery average _____ and find the residual, if in fact valve surgery averaged 34 thousand dollars at Allegheny General. _____
- (f) The regression output information suggests that the size of your prediction error in (e) is (i) unusually small (ii) unusually large (iii) fairly typical

The regression equation is

$$\text{Valve} = -1.7 + 1.50 \text{ Bypass}$$

$$S = 4.35$$

$$R\text{-Sq} = 80.2 \%$$

$$R\text{-Sq}(\text{adj}) = 77.7 \%$$

